

Changes in Diagnostic Decision-making After a Computerized Decision Support Consultation Based on Perceptions of Need and Helpfulness: A Preliminary Report

Fredric M. Wolf, Ph.D.^a, Charles P. Friedman, Ph.D.^b, Arthur S. Elstein, Ph.D.^c, Judith G. Miller^d, Gwendolyn C. Murphy, Ph.D.^e, Paul Heckerling, M.D.^c, Paul Fine, M.D.^d, Tom Miller, M.D.^e, James Sisson, M.D.^d, Sema Barlas, Ph.D.^c, Amy Capitano^d, Macy Ng^c, Tim Franz^c

^aUniversity of Washington, Seattle

^bUniversity of Pittsburgh, Pittsburgh

^cUniversity of Illinois, Chicago

^dUniversity of Michigan, Ann Arbor

^eUniversity of North Carolina, Chapel Hill

We examined the degree to which attending physicians, residents, and medical students' stated desire for a consultation on difficult-to-diagnose patient cases is related to changes in their diagnostic judgments after a computer consultation, and whether, in fact, their perceptions of the usefulness of these consultations are related to these changes. The decision support system (DSS) used in this study was ILIAD (v4.2). Preliminary findings based on 16 subjects' (6 general internists, 4 second-year residents in internal medicine, and 6 fourth-year medical students) workup of 136 patient cases indicated no significant main effects for 1) level of experience, 2) whether or not subjects indicated they would seek a diagnostic consultation before using the DSS, or 3) whether or not they found the DSS consultation in fact to be helpful in arriving at a diagnosis ($p > .49$ in all instances). Nor were there any significant interactions. Findings were similar using subjects or cases as the unit of analysis. It is possible that what may appear to be counter-intuitive, and perhaps irrational, may not necessarily be so. We are currently examining potential explanatory hypotheses in our ongoing current, larger study.

INTRODUCTION

There is growing interest in evaluating in a more rigorous fashion the utility of the increasing number of decision support tools that are becoming available in medicine. Taking cases as the unit of analysis, we had statistical Power of 80% or greater to detect an overall change in the differential diagnosis from pre to post DSS consultation of 15% or more at $p < .05$, two-tailed.¹⁻³ In a "consumers report" type analysis by Berner and colleagues⁴ in which the study investigators evaluated how four diagnostic decision support systems performed on an identical set of 105 diagnostically challenging cases, performance (diagnostic accuracy) ranged between 52% and 71%, with no system clearly superior in all areas. On the basis of this analysis, Kassirer⁵ gave these tools a

mixed review, stating they had essentially achieved a grade of "C". In an ideal world, it would be desirable for any consultation, human or otherwise, to receive a grade of "A", indicative of a high degree of diagnostic accuracy, perhaps 90% or greater. The question remains, how good is good enough for these types of tools to gain widespread acceptance and use?

A multicenter collaborative study has been designed to examine the usefulness of consulting such decision support systems (DSS). Preliminary results reported by Elstein and colleagues⁶⁻⁷ suggest that such DSS consultations affect diagnostic judgments in a minority of cases, even when the cases are extremely difficult to diagnose. Moreover, Elstein et al observed that their data suggest that many clinicians report that they perceive the DDS to be more useful than their actual decision making behavior would suggest. The present report examines the degree to which attending physicians, residents, and medical students' stated desire for a consultation on difficult-to-diagnose patient cases is related to changes in their diagnostic judgments after a computer consultation, and whether, in fact, their perceptions of the utility of these consultations are related to these changes. The two research hypotheses tested are that attending physicians, medical residents, and medical students are more likely to change their initial differential diagnosis after using a DSS when they 1) indicate the patient case is one for which they would actually seek an outside opinion or consultation, and 2) they find the DSS consultation to be helpful. In addition, differences related to these two hypotheses among the three groups of subjects based on their level of medical experience were examined, although no specific interaction effects are hypothesized.

METHODS

Overview

Eighteen subjects, 6 general internists, 6 second-year residents in internal medicine, and 6 fourth-year medical students were recruited to participate in a

pilot study to test the methods and design for a larger multicenter study to evaluate the impact of several computerized diagnostic decision support systems on clinical judgments. One-third of the sample of subjects in each of the three "levels of experience" groups were recruited from each of three Medical Schools (the Universities of Illinois, North Carolina, and Michigan). Two residents (one from each of two sites) failed to complete the protocol, thus leaving a total of 16 subjects in the study. Each subject did a diagnostic workup for a set of 9 diagnostically difficult cases (as described below) before and after using the computerized DSS, thereby providing a sample of 144 cases of which 136 were available for analysis after eliminating cases with missing data. The DSS used in this study was ILIAD 4.2, a combined Bayesian and Boolean based expert diagnostic system for internal medicine containing over 2400 diseases and medical syndromes, and over 11,000 findings in its knowledge base.⁸

Case Materials

A set of 36 diagnostically challenging cases were selected from recent admissions to internal medicine services, 12 from each of the three University teaching hospitals. All cases had discharge diagnoses listed in ILIAD's knowledge base. A 2-3 page CPC-type case summary was prepared for each case. These included history, physical findings, and laboratory tests in the patient's chart except for the definitive diagnostic information (typically but not always a laboratory test), which would make the case unchallenging and obviate the need for a consultation, human or otherwise. This is similar to the procedure described in the Berner et al study.⁴ All case summaries were rated for difficulty by the clinician investigators, one at each site. Only cases rated 3 or greater on a 7-point difficulty scale by all 3 raters were retained for the study. To design a task of manageable length for each subject, a matrix sampling approach was used in which the 36 cases were randomly assigned to 1 of 4 clusters of 9 cases in a stratified fashion to balance each cluster (set) for difficulty and site of origin. This approach is intended to improve the generalizability of the findings back to a larger domain of medical problems.

Design and Procedures

Subjects were trained to use ILIAD with a standardized training protocol used across the 3 sites. The training protocol focused on the major features of entering data, consulting disease frames, and interpreting the DSS's differential diagnosis list. Subjects were randomly assigned to a cluster. Each subject worked up a cluster of 9 cases, first without and then with an ILIAD consultation. To avoid possible effects of order of presentation of cases within a cluster, the order was randomized for each subject, except that the first two cases administered were always the easiest of the set. For each case,

they were instructed to offer a differential diagnosis list of up to six diagnoses and to indicate how likely they were to seek a diagnostic consultation on a four-point scale (definitely/probably/probably not/almost certainly not). Subjects then entered whatever findings they chose to into the DSS. After concluding the DSS consultation, they were again asked to provide their differential diagnosis, as well as their ratings of the helpfulness of the DSS on a 4-point scale (very helpful/generally helpful/slightly helpful/not helpful).

For purposes of analysis, the 4-point "consultation" and "helpfulness" scales were collapsed into dichotomous scales where "definitely" and "probably" seek an outside consultation to help make a diagnosis were both coded as "yes" and "probably not" and "almost certainly not" coded as "no". Similarly, for the helpfulness scale, "very helpful" and "generally helpful" were coded as "yes" and "slightly" or "not at all helpful" were coded as "no."

Statistical Analysis

The three independent variables were 1) need to seek a diagnostic consultation, 2) perceived helpfulness of the DSS consultation, and 3) level of medical experience. The dependent variable was whether or not the position of the correct diagnosis for each case changed as a result of using the DSS. When the correct diagnosis did not appear on either the subject's pre- or post-DSS consultation list, it was scored as "no change." A 2 x 2 x 2 contingency table analysis⁹ was performed to test the hypotheses and research questions. In order to facilitate interpretation of the results, two separate 2 x 2 contingency table analyses were also performed, one for "seek consultation" x experience and the second for "helpfulness" x experience. Analyses were performed with both subject (n=16) and cases (n=136) as the unit of analysis.

RESULTS

Findings were similar using subjects or cases as the unit of analysis. For ease of presentation, only case results are reported here (see Table 1). There were no significant main effects for 1) level of experience, 2) whether or not subjects indicated they would seek a diagnostic consultation before using the DSS, or 3) whether or not they found the DSS consultation in fact to be helpful in arriving at a diagnosis ($p > .49$ in all instances). Nor were there any significant 2- or 3-way interactions ($p > .21$ in all instances). Visual summaries of the findings may be seen in Figures 1 and 2. Results were consistent with the direction of research hypothesis 1, although not statistically significant, for medical students and residents (Figure 1). That is, they tended to change their differential diagnosis in a slightly, but not significantly, greater number of cases when they indicated they would

(rather than would not) seek an outside diagnostic consultation. Attending physicians, on the other hand, indicated just the opposite. Results were similarly consistent with the direction of research hypothesis 2, although not statistically significant, in the case of medical students and attending physicians,

but not residents (Figure 2). Students and attendings tended to change their differential diagnosis in a slightly, but not significantly, greater number of cases when they indicated the DSS consultation was helpful (rather than not helpful). Residents, on the other hand, indicated just the opposite.

Table 1. Percentage (%) of 136 patient cases in which 6 medical students, 4 residents, and 6 attending physicians in General Medicine changed their diagnosis on the basis of whether they would seek an outside diagnostic consultation before using a computerized decision support system and whether they believed the consultation helpful or not after using it.

Group	Consultation		Helpful	
	Seek	Not Seek	Yes	No
Medical students	29.6	28.6	32.0	27.6
Residents	17.2	14.3	8.3	20.8
Attendings	11.5	25.0	30.0	15.4

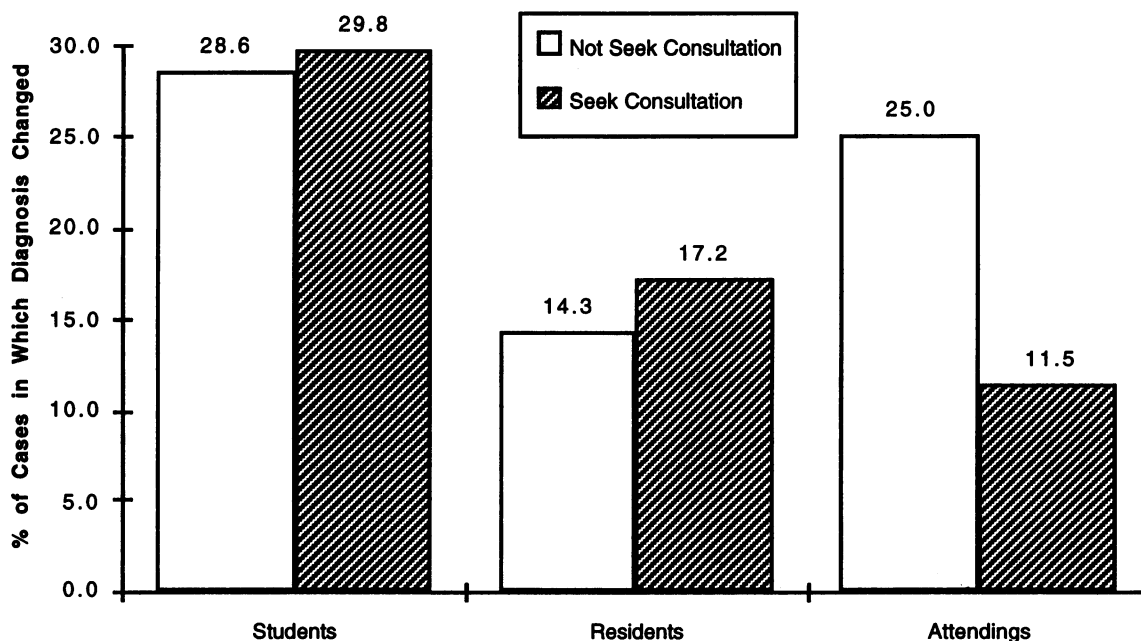


Figure 1. Percentage of 136 patient cases in which 6 fourth-year medical students, 4 second-year residents, and 6 attending physicians in General Medicine changed their diagnosis on the basis of whether they would seek an outside diagnostic consultation before using a computerized decision support system.

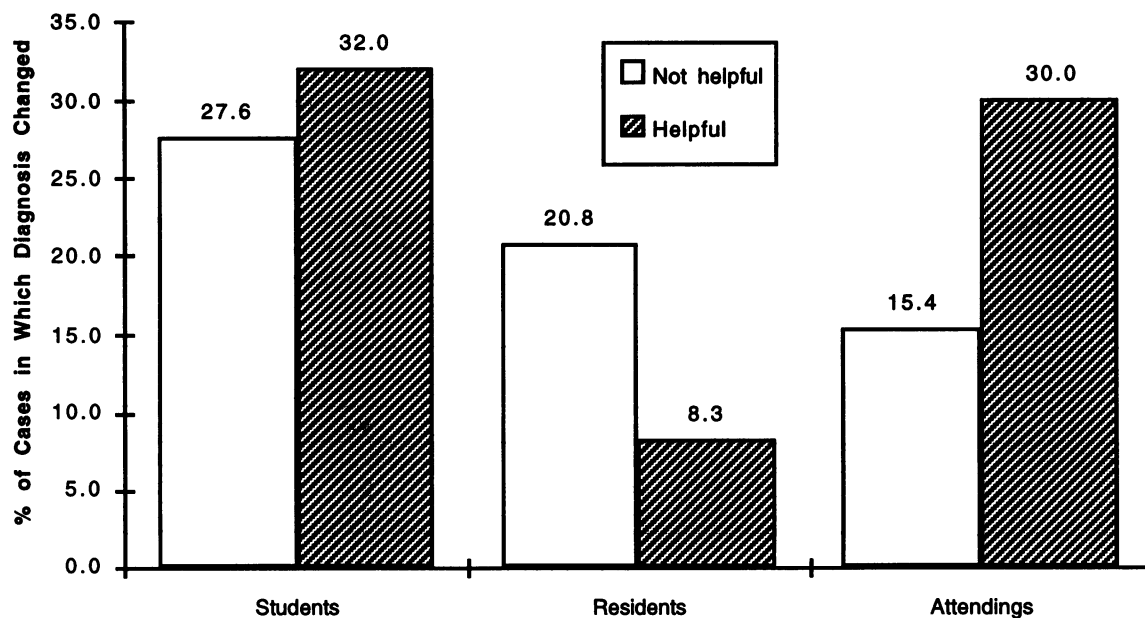


Figure 2. Percentage of 136 patient cases in which 6 fourth-year medical students, 4 second-year residents, and 6 attending physicians in General Medicine changed their diagnosis on the basis of whether they found a computerized decision support system diagnostic consultation to be helpful.

DISCUSSION

These preliminary results from the pilot study did not provide support for either of the research hypotheses. Whether or not subjects thought they would seek an outside diagnostic consultation in advance of using a computerized decision support system, and whether or not they thought the DSS consultation in fact was helpful after using the system, did not influence whether or not they changed their differential diagnosis after consulting the system. It is possible, however, that these findings may change when more subjects and cases are examined, or when other outcomes, such as certainty or confidence in the diagnosis are considered. Taking cases as the unit of analysis, we had statistical Power of 80% or greater to detect an overall change in the differential diagnosis from pre to post DSS consultation of 15% or more at $p < .05$, two-tailed.¹⁰ Some of the observed differences did approach 15% and may perhaps reach statistical significance in the larger, ongoing study.

It is possible that subjects (particularly attendings) may logically not change their differential diagnosis even when they think a consultation might be necessary because the consultation may in fact have not been helpful, or alternatively, may have reinforced their initial diagnostic judgments. Thus the DSS may have served to help confirm their initial impressions, thereby being perceived as useful even

though they may not have changed their diagnostic judgments. Similarly, it is possible that subjects may logically not change their differential diagnosis even when they think the DSS consultation was helpful because the consultation may in fact have reinforced their initial differential. We are currently examining these potential explanatory hypotheses in our ongoing current, larger study. The addition of a measure of diagnostic confidence and certainty in the larger study may help provide some explanatory rationale.

Although there was a substantial number of patient cases used in the present study, the sample of students, residents, and attendings was modest at best. Thus it is always possible that these results may not be entirely representative of the larger population of peers whom our subjects are intended to represent, although there is no readily apparent reason why such bias should exist.

Acknowledgments

This work was supported in part by a grant from the National Library of Medicine, Grant number NIH 1 R01 LM 05630. We would like to thank David Potts, Kevin Biolsi, Xiao Mei, Keith Cogdill, Amina Barnes, and Alice Nabalamba for their assistance with the project.

Correspondence may be directed to Fredric M. Wolf, Ph.D., Department of Medical Education, University of Washington, E-312 Health Sciences/Box 357240, Seattle, WA 98195-7240.
E-mail: wolf@u.washington.edu

References

1. Johnston ME, Langton KB, Haynes RB, Mathieu A. Effects of computer-based clinical decision support systems on clinical performance and patient outcomes: a critical appraisal of research. *Ann Intern Med* 1994; 120:135-142.
2. Miller RA. Medical diagnostic decision support systems--past, present and future: a threaded bibliography and brief commentary. *JAMIA* 1994;1:8-27.
3. Workshop on Evaluation of Knowledge-based Systems, December 7-8, 1995. Bethesda, MD: National Library of Medicine/AHCPR/NSF.
4. Berner ES, Webster GD, Shugerman AA, et al. Performance of four computer-based diagnostic systems. *N Engl J Med* 1994; 330:1792-6.
5. Kassirer JP. A report card on computer-assisted diagnosis--the grade C. *N Engl J Med* 1994; 330:1824-5.
6. Elstein AS, Friedman CP, Wolf FM, et al. Evaluating a decision support system (abstract). Invited paper presented at the Workshop on Evaluation of Knowledge-based Systems, December 7-8, 1995. Bethesda, MD: National Library of Medicine/AHCPR/NSF.
7. Elstein AS, Friedman CP, Wolf FM, et al. Effects of a decision support system on the diagnostic accuracy of users: a preliminary report. *JAMIA* 1996;3: 422-428.
8. Applied Medical Informatics. *ILIAD v4.2: Physician Diagnostic and Treatment Software*. Salt Lake City: Applied Medical Informatics, 1994.
9. Conover WJ. *Practical Nonparametric Statistics*, second ed. New York: Wiley, 1980.
10. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, second edition. Hillsdale, NJ: Lawrence Erlbaum, 1988.